

Message

From: Guy Rosen [/O=THEFACEBOOK/OU=EXTERNAL
(FYDIBOHF25SPDLT)/CN=RECIPIENTS/CN=]
Sent: 2/12/2019 9:14:03 AM
To: hpm-integrity [hpm-integrity@fb.com]
BCC: Christina [redacted]@fb.com]; Michael Ivan [redacted]@fb.com]; Ashwini [redacted]
[redacted]@fb.com]; James [redacted]@fb.com]; Will [redacted]@fb.com]; Alex [redacted]@fb.com];
James [redacted]@fb.com]; Andrew Bocking [redacted]@fb.com]; Guillermo [redacted]@fb.com]; Ruben
[redacted]@fb.com]; Robert [redacted]@fb.com]; Taichi [redacted]
[redacted]@fb.com]; Tian [redacted]@fb.com]; Ollie [redacted]@fb.com]; Bo [redacted]@fb.com]; Sammie [redacted]
[redacted]@fb.com]; William [redacted]@fb.com]; Zeev [redacted]@fb.com]; Stan Chudnovsky
[redacted]@fb.com]; Stephane [redacted]@fb.com]; Laurent [redacted]@fb.com]; Tony [redacted]@fb.com];
Ted [redacted]@fb.com]; Drew [redacted]@fb.com]; Tracy [redacted]@fb.com]; Michelle [redacted]
[redacted]@fb.com]; Sam [redacted]@fb.com]; Kris N [redacted]@fb.com]; Sheridan
[redacted]@fb.com]; Mitali [redacted]@fb.com]; Devdatta [redacted]@fb.com]; Peter [redacted]
[redacted]@fb.com]; Mike [redacted]@fb.com]; Jonathan [redacted]@fb.com]; Ezequiel [redacted]
[redacted]@fb.com]; Michael [redacted]@fb.com]; Jonathan [redacted]@fb.com]; David [redacted]@fb.com];
Glen [redacted]@fb.com]; Jessica [redacted]@fb.com]; Kate Gotimer [redacted]@fb.com]; Mick
[redacted]@fb.com]; Nivedita [redacted]@fb.com]; Nidhi [redacted]@fb.com];
Thomas [redacted]@fb.com]; Angela [redacted]@fb.com]; Dana [redacted]@fb.com];
Stephen [redacted]@fb.com]; Chris [redacted]@fb.com]; Shirley [redacted]@fb.com]; Lizzy [redacted]
[redacted]@fb.com]; Erez [redacted]@fb.com]; Kelly [redacted]@fb.com]; Ted [redacted]@fb.com]; Sagee
[redacted]@fb.com]; Jon [redacted]@fb.com]; Aaron [redacted]@fb.com]; Larry [redacted]
[redacted]@fb.com]; Bertram [redacted]@fb.com]; Alex [redacted]@fb.com]; Valerio [redacted]@fb.com];
Mike [redacted]@fb.com]; Mary [redacted]@fb.com]; Al [redacted]@fb.com]; Julien [redacted]
[redacted]@fb.com]; Neha [redacted]@fb.com]; Vlad [redacted]@fb.com]; Amin [redacted]
[redacted]@fb.com]; Cristian [redacted]@fb.com]; Shuo [redacted]@fb.com]; Sharon [redacted]
[redacted]@fb.com]; Ben [redacted]@fb.com]; Yuan [redacted]@fb.com]; Stefan [redacted]@fb.com];
Patrick [redacted]@fb.com]; Ryan [redacted]@fb.com]; Nicola [redacted]@fb.com];
Guillermo [redacted]@fb.com]; Iris [redacted]@fb.com]; David [redacted]@fb.com]; Alicia
[redacted]@fb.com]; Pratiti Raychoudhury [redacted]@fb.com]; Brian [redacted]@fb.com];
Jen [redacted]@fb.com]; Deborah [redacted]@fb.com]; Will Cathcart [redacted]@fb.com]; Ami [redacted]
[redacted]@whatsapp.com]; Fidji Simo [redacted]@fb.com]; Dan [redacted]@fb.com]; Rob [redacted]
[redacted]@fb.com]; John Hegeman [redacted]@fb.com]; Claudia [redacted]@fb.com]; Oleg [redacted]
[redacted]@fb.com]; Ber [redacted]@fb.com]; Y [redacted]@fb.com]; Kevin [redacted]@fb.com]; Ginger Baker
[redacted]@fb.com]; Lindsey [redacted]@fb.com]; James [redacted]@fb.com]; Marija [redacted]
[redacted]@fb.com]; Tanton [redacted]@fb.com]; Ying-Yi [redacted]@fb.com]; Li-ye [redacted]@fb.com];
Kenny [redacted]@fb.com]; Ian [redacted]@fb.com]; Ivan [redacted]@fb.com]; Daniel [redacted]
[redacted]@fb.com]; Omri [redacted]@fb.com]; Christopher [redacted]@fb.com]; Kevin [redacted]
[redacted]@fb.com]; Jorge [redacted]@fb.com]; Vasanth [redacted]@fb.com]; Victor [redacted]
[redacted]@fb.com]; Goran [redacted]@fb.com]; John [redacted]@fb.com]; Isaac [redacted]@fb.com];
Hervé [redacted]@fb.com]; Jieqi [redacted]@fb.com]; Shenghao [redacted]@fb.com]; Volodymyr [redacted]
[redacted]@fb.com]; Angeline [redacted]@fb.com]; Shoeb [redacted]@fb.com]; Sagar [redacted]@fb.com]; Tim
[redacted]@fb.com]; Adam Mosseri [redacted]@instagram.com]; Greg [redacted]@fb.com];
William [redacted]@fb.com]; Naomi [redacted]@fb.com]; Ridthi Sanjanwala [ridthi@fb.com]; Sam [redacted]
[redacted]@fb.com]; Kelly [redacted]@fb.com]; Terri [redacted]@fb.com]; Ellen [redacted]
[redacted]@fb.com]; Luke [redacted]@fb.com]; Guy Rosen [redacted]@fb.com]; Naomi Gleit [redacted]@fb.com]; David
[redacted]@fb.com]; Chad [redacted]@fb.com]; Annette [redacted]@fb.com]; Ziqing [redacted]
[redacted]@fb.com]; Ross [redacted]@fb.com]; Justin Osofsky [redacted]@fb.com]; David Gillis [redacted]@fb.com];
Elisabeth [redacted]@fb.com]; Monika Bickert [redacted]@fb.com]; Danny Ferrante [redacted]@fb.com]; Brady
Lauback [redacted]@fb.com]; Alex Schultz [redacted]@fb.com]; Javier Olivan [redacted]@fb.com]; Ashley [redacted]
[redacted]@fb.com]; Denise [redacted]@fb.com]; Irina [redacted]@fb.com]; Jasmine [redacted]
[redacted]@fb.com]; Veronika Belokhvostova [redacted]@fb.com]; Betty [redacted]@fb.com]; Fariba
[redacted]@fb.com]; Helaina [redacted]@fb.com]; Natalia [redacted]@fb.com];
Abigail Sooy [redacted]@fb.com]; Sev [redacted]@fb.com]; Scott [redacted]@fb.com]; Alex [redacted]
[redacted]@fb.com]; Maggie Louise [redacted]@fb.com]; Ilona [redacted]@fb.com]; Gearoid [redacted]
[redacted]@fb.com]; Kevin [redacted]@fb.com]; Brandon [redacted]@fb.com]; Shauna [redacted]
[redacted]@fb.com]; Jamie [redacted]@fb.com]; Laura [redacted]@fb.com]; Mirelle [redacted]

CONFIDENTIAL
CONFIDENTIAL

META3047MDL_REPROD_VOL025
METANMAG_REPROD_VOL018

META3047MDL-003-00144080
METANMAG-002-00144621
PLTF EX-00370_0001

PLTF EX

PLTF EX-00370

[redacted]@fb.com]; Andrea Sau [redacted]@fb.com]; Tammy [redacted]@fb.com]; Mike [redacted]
[redacted]@fb.com]; Laura [redacted]@fb.com]; Cody [redacted]@fb.com]; Bradley [redacted]
[redacted]@fb.com]; Alex [redacted]@fb.com]; Yvonne [redacted]@fb.com]; Richard [redacted]
[redacted]@fb.com]; Matt [redacted]@fb.com]; Raychel [redacted]@fb.com]; Rache [redacted]
[redacted]@fb.com]; Tiffany [redacted]@fb.com]; Adam [redacted]@fb.com]; Sherhan [redacted]
[redacted]@fb.com]; Marco [redacted]@fb.com]; Nathan [redacted]@fb.com]; Amit [redacted]
[redacted]@fb.com]; Jeff King [redacted]@fb.com]; Olga [redacted]@fb.com]; Leah [redacted]
[redacted]@fb.com]; Carmen [redacted]@fb.com]; Alberto [redacted]@fb.com]; Moayad [redacted]
[redacted]@fb.com]; Constantine [redacted]@fb.com]; Matthew [redacted]@fb.com]; Melanie [redacted]
[redacted]@fb.com]; Lisa [redacted]@fb.com]; Jacqueline [redacted]@fb.com]; [redacted]
[redacted] [kvoekery@fb.com]; Sanchi [redacted]@fb.com]; Madelyn [redacted]@fb.com]; Eli [redacted]
[redacted]@fb.com]; Monica [redacted]@fb.com]; Lauren [redacted]@fb.com];
Assaf [redacted]@fb.com]; Melanie [redacted]@fb.com]; Zoe [redacted]@fb.com]; Luis Fernando [redacted]
[redacted]@fb.com]; Craig [redacted]@fb.com]; Jen [redacted]@fb.com]; Jay [redacted]
[redacted]3@fb.com]; Susan [redacted]@fb.com]; Ryan [redacted]@fb.com]; Liliya [redacted]
[redacted]@fb.com]; PJ [redacted]@fb.com]; Philip [redacted]@fb.com]; Brandi [redacted]
[redacted]@fb.com]; Navila [redacted]@fb.com]; Luca [redacted]@fb.com]; Briana [redacted]
[redacted]@fb.com]; Anne [redacted]@fb.com]; Emmanuel [redacted]@fb.com]; Emily [redacted]
[redacted]@fb.com]; Declan [redacted]@fb.com]; Rita [redacted]@fb.com]; Leslie [redacted]
(eCrime) [redacted]@fb.com]; Ryan [redacted]@fb.com]; Meagan [redacted]@fb.com]; Karine [redacted]
[redacted]@fb.com]; Elizabeth [redacted]@fb.com]; Nektarios [redacted]@fb.com];
John [redacted]@fb.com]; Umer Farooq [redacted]@fb.com]; Eleni [redacted]@fb.com]; Asim [redacted]
[redacted]@fb.com]; Eoghan [redacted]@fb.com]; Anh [redacted]@fb.com]; Catherine [redacted]
[redacted]@fb.com]; April [redacted]@fb.com]; Jen [redacted]@fb.com]; Jason [redacted]
[redacted]@fb.com]; Timothy [redacted]@fb.com]; Ryan [redacted]@fb.com]; Yinyin [redacted] (not
Yinyi) [redacted]@fb.com]; Cynthia [redacted]@fb.com]; William [redacted]@fb.com]; Bill [redacted]
[redacted]@fb.com]; Maggie [redacted]@fb.com]; Kevin Weil [redacted]@fb.com]

Subject: Integrity HPM - Feb 11 2019

SUMMARY

It's been a while since my last Integrity HPM, this marks the first of this year and I plan to keep it going. H1 is underway full speed ahead, with teams all over the first priority of our company's 2019 plan: *continue making progress on the major social issues facing the internet and our company*. As laid out in our Integrity H1/H2 review ([readout](#)) the strategy for the year falls into:

1. **Get the high-priority problems under control** across surfaces
2. **Get buy-in** from people and from stakeholders
3. **Build sustainably** through centralization, common platforms and efficiency

We also have a few cross-cutting themes, such as our continued work on elections and critical countries (more details below), advancing our work on content enforcement including improving accuracy ([priorities here](#)) and getting ready for e2e encryption (more below).

Prioritization

HIGHLIGHTS

Rollouts underway for our tight election calendar: India, Indonesia and EU are our top priority elections

(Civic Integrity, Business Integrity, FAST Integrity, Pages Integrity, Community Integrity, Integrity Investigations & Intelligence, WhatsApp Integrity, many others...)

- Coming out of Brazilian elections and the US midterms at the end of 2018, our biggest challenge is to productize our election efforts so that we have the right defenses scaled globally. This is a lot of work, which will take time. Which means in the meantime we have to ruthlessly prioritize: on the one hand, much of the custom support this half will go towards just three top-priority elections: IN, ID, EU; on the other hand, most of the remaining energy will go towards building scaled solutions. Some of these will be ready to roll out in additional countries,

and some will not. This isn't easy: H1 is a particularly tight half with many elections around the world:
<https://our.internmc.facebook.com/intern/elections/>

- Political ads:
 - India – last week we started to enforce authenticity of electoral ads and launched the ad library; in Nigeria we started to reject all foreign electoral ads, a temporary measure where we're unable to launch authenticity measures in time.
 - In Thailand and Indonesia, we plan to launch similar measures to reject foreign electoral ads
 - In the EU, we will enforce authenticity of electoral and issue ads
 - In Ukraine and Israel we will enforce authenticity of electoral ads
 - The team is working around the clock on a super tight schedule to get these ready. The thing with election dates is... they do not move! We hope to have a global platform to enforce authenticity of electoral ads by the end of H1.
- Misinformation:
 - In Nigeria, we've expanded our 3PFC program to three fact checkers (Dubawa, AFP Nigera, Africa Check), and launched temporary measures to increase our defenses against potential misinformation, lowering the threshold for our native content soft demotion.
 - We're working to launch Chain of Trust, our crowdsourced-based fact checking program ahead of the India Indonesia elections. Chain of Trust brings vastly improved volumes and lower latency as compared to 3PFC, and we're looking forward to go live. We're also working to kick off a more collaborative public process around Chain of Trust in order to build more support and make the program more independent in the long term.
 - We launched fact-checking in Arabic, in partnership with Agence France-Presse (AFP), that will have impact throughout Northern Africa and the Middle East.
- Info Ops:
 - We're executing at a brisk pace of takedowns. Just last year each and every takedown was a massive lift and endless leadership reviews. The team has set up streamlined processes, from investigative reviews to takedown approvals, which has enabled a step change in our pace (also, we've hired lots of people).
 - In the last two weeks, we've taken down clusters related to [Indonesia](#) and [Iran](#). Since the beginning of the year, we've also taken down a [domestic activity in the Philippines](#) and [Russian activity](#) targeting the Baltics.
 - Quote of the week from [The Verge](#): *"the trend is clear around the world: Facebook is enforcing its policies around coordinated inauthentic behavior and incitement to violence in more places, and more publicly, than it has to date."*
- The election war room is now called the Integrity Product Operations Center (IPOC), and we're working to bring new locations in MPK, Singapore and Dublin online ahead of the coming elections. More updates here: <https://fb.workplace.com/groups/799496770098940/permalink/2316982641683671/>

Reporter appeals rolled out to all problem types

(Community Integrity, Community Operations, Product Policy)

- We have two flavors of appeals on content decisions: one usually thinks of actor appeals, which is when you appeal our decision to take down something you posted. But, also important is reporter appeals, which is when you appeal our decision to NOT take down something you reported. Reporter appeals are especially important for safety-related areas, where we really want a way for people to tell us "seriously, take a look".
- Over the past year, we've rolled out actor appeals for most problem type, and we just flipped the switch on reporter appeals across the board. This is an important milestone, and also unlocks more flexibility on our ability to in the automate inbound user reports, given we provide an option for people to appeal our initial decision.
- The next frontier with our takedown and appeals experiences is rebuilding our penalty system, and instrumenting "procedural fairness" – a key part of our "buy in" pillar to help us understand how we build an experience that provides transparency, voice and fairness. Our goal is to be able to run tests to prove that we can move procedural fairness by creating a better experience.

Instagram suicide and self-injury (SSI) issues reveal we are behind where we need to be
(Instagram Wellbeing, Community Integrity, Community Operations, Product Policy, Comms)

- We've always removed content that promotes SSI, but at the advice of mental health experts, we've allowed suicide admission content as it provides an outlet for people to cope and get support. Following the tragic suicide of a 14-year-old girl in the UK that her family linked to her access to self-harm content, Instagram has been under intense scrutiny over the last few weeks around our policies and enforcement.
- Policy: after further consultation, we decided that graphic cutting images should be removed as the potential for modeling, normalizing, or triggering self-harm in others outweighs the value of sharing the content. Non-graphic self-mutilation or healed cuts imagery will be left up but blurred / marked as sensitive and removed from discovery.
- Enforcement: because content outside of promotion did not violate our policies, we had no detection or enforcement in place at all for any other content, and most worryingly the content was easy to discover through search, hashtags, IG Explore, etc. We have a swat team across IG Wellbeing, Community Integrity and Community Operations sprinting to clean up hashtags, clean up content, apply warning screens, and reduce visibility of potential SSI content from discovery surfaces.
- We also really need to build proactive detection for content. This means ensuring we have the review flows set up right to label the content we're finding, and training our classifiers. We're optimistic we have the technology to build computer vision detection for this space, it will just take work and time to get the classifiers working well, and in the meantime given we've announced our new policy it will be too easy to find content we've left up.
- A big callout to the local comms and policy teams who faced a very negative cycle, and Adam who flew to the UK to speak to NGOs, the government and press. We're certainly not out of the woods – we have good plans but it will take time to deliver on what we need.
- To me, this is a painful lesson on what we mean when we say areas are at “risk” in our maturity framework. Why was SSI in this state? As some folks may remember, in 2017 we had a lockdown around FB Live in particular, focusing on getting help to people expressing suicidal thoughts. The work we did then has resulted in hundreds of wellness checks from first responders that has saved lives. In retrospect, that good work gave us a false sense of security on our SSI work, and as we ramped up our teams we chose to first put people first on other critical areas that sorely needed attention, such as child safety, terrorism and drugs. This half we rebooted our SSI efforts, and literally days before this became a full-blown fire, we were in the first roadmap review for the new CI SSI team – fittingly, the key takeaway was that the space is broader and more complex than the “imminent harm” work we'd historically done. I'm glad we have a team ready to jump in, but it's painful to see the gap we left.

Buy-in: kicked off consultation for the independent board, talking to the French regulatory working group
(Policy, Comms, Community Integrity, Community Operations, Marketing)

- Progress on two key projects in our “stakeholder buy-in” pillar:
- First, the independent oversight board:
 - We announced we'd pursue this in Mark's November letter, and a couple of weeks ago we announced (newsroom post + Nick at Brussels) the draft charter for the board.
 - One of the key realizations we had late last year is that if we're going to create a body that takes decision-making power from the company, it would be ironic if we decided everything about how it works. Therefore, the draft charter marks the start of a consultation process that includes a number of events around the world and lays out questions, not answers, about how the board could work. Since there are no right answers here, it's incredibly important to engage with the outside world, to listen and to build support over time.
 - This first announcement landed neutral to positive, with quotes from influencers such as “Does Facebook's Oversight Board charter raise questions about independence and representation? Sure. Is it better than what we've got now? Probably”, “A rare and fascinating devolution of power”, “They started

this process in an admirable way, and we should help ensure it really does have a positive impact". Much of this is due to excellent execution on our comms and policy teams that worked to ensure even this first announcement would not be a "big reveal" but instead operated a rigorous process to stay in constant touch with thought leaders ahead of this milestone. We hope to continue this style of execution as we go into the next phase.

- Second, the regulatory working group with the French government
 - Last November, we kicked off a six month trial with the French government where they will take the time to go deep and truly understand our policies, processes and systems as they think about designing future regulation (<https://techcrunch.com/2018/11/12/facebook-to-let-french-regulators-investigate-on-moderation-processes/>)
 - The French government put together an impressive, forward-thinking group with representatives of regulators, judiciary, key government departments and technical folks such as computer science professors, the government CTO and others. They've been very engaged in the process.
 - So far we've held 5 deep-dive sessions to educate on our business, our policies and how they are updated, our enforcement system (including a visit on-site to a CO review site), AI, metrics and tools (including a shadowing session on real but redacted content), etc. We've made an effort to be open and transparent with the group.
 - I was personally in Paris two weeks ago to participate in the session on our systems and metrics. I found the group thoughtful, constructive and quite sharp. I even drew a precision & recall curve on the whiteboard and had a thoughtful discussion! (And they understood it well, making the observation that you have to pick one: you can't take all the bad content down without false positives, or avoid false positives without leaving bad content up). Needless to say we don't see this kind of depth too often.
 - We have largely concluded the education phase of our work, with some smaller follow ups. The French team is now engaging outside of Facebook, including meeting the German government to understand their regulatory approach (the NetzDG law). We are hopeful that the time they took and our efforts will help design regulation that appropriately incentivizes us to improve our systems.
 - This process, as well as many recent cases where we invited outside experts into our meetings or review sites, has built up an appetite to make this more programmatic, creating a compressed curriculum that we can use with other governmental, academic or other bodies. We consistently find that once people are brought into the fold, they realize the complexity and recognize our thoughtful approach. Scaling this could be very impactful.

Work underway to understand impact of e2e encryption and approach to safety

(Messenger Integrity, Community Integrity)

- End-to-end encryption is a key part of the company's strategic approach to messaging. This is one of those hard cases where privacy and safety can be in tension, so we have to understand our new approach to enforcement in this future state. (More context in [this post from last year.](#)). Work is happening across Messenger and CI teams.
- First, the teams are working to understand how to roll out e2ee responsibly in the messaging product. For example, while we don't want a state where encryption can be turned off by a server-side flag based on some arbitrary attributes, we are exploring if there could be different, clearly delineated product surfaces (e.g., the requests tab) where encryption might not be active.
- The CI Problem teams most dependent on messaging access are dedicating at least 50% of their roadmaps to understanding this future state. We're testing the effectiveness of behavioral-only detection, understanding the upstream discovery funnel where we will need to put mitigations (making it harder for attackers to find victims), and exploring how to aggressively upsell report flows. One promising early example from the scams team showed that withholding content signals from our labeling flow reduced classifier positive rates by just 1%. Scams are very much behavioral, and while we may see similar results in other cases where attackers use patterns of behavior to target large numbers of victims, it's the smaller scale cases where relationships are built over time that will likely prove much harder to tackle.

- To better understand our “unknown unknowns”, we're also working towards “SieveStorm” – an exercise where we will turn off message access at a low level, in order to identify what other systems, processes and teams are going to break with e2ee. Planning is underway.

HPMS FROM THE TEAMS

- [Community Integrity](#)
- [Civic Integrity](#)
- [Business Integrity](#)
- [FAST Integrity](#)
- [Integrity Infrastructure](#)
- [Instagram Wellbeing](#)
- [Video Integrity](#)
- Messenger Integrity
- [Pages Integrity](#)
- [Groups Integrity](#)
- Marketplace Integrity

ME

I spent the winter break with family and friends in Costa Rica, where I learned about Pura Vida. Literally “pure life”, Pura Vida is a Costa Rican “tagline” that has grown over the past few decades to be used for hello, goodbye, everything's great, everything's cool, no worries – it reflects the country's laid back culture, no stress, no fuss, be thankful for what you have and don't dwell on the negative.

The last months of 2018 at work felt honestly kind of brutal, but the vacation really filled up my reserves (also, y'know, I slept!). I returned feeling ready to go and more resilient than ever. So if I still seem cool and calm despite another fire, impossible prioritization and the intimidating mountain of work ahead of us – it's that Pura Vida spirit!